Received: 26 February 2019 | Revised: 1 September 2019 | Accepted: 17 October 2019
DOI: 10.1002/asi.24321

RESEARCH ARTICLE

JASIST WILEY

Co-contributorship Network and Division of Labor in Individual Scientific Collaborations

Chao Lu[1] | **Yingyi Zhang**[2] | **Yong-Yeol Ahn**[3] | **Ying Ding**[4] | **Chenwei Zhang**[3,5] | **Dandan Ma**[6]

[1]Business School, Hohai University, Nanjing, Jiangsu, China
[2]School of Economics and Management, Nanjing University of Science and Technology, Nanjing, Jiangsu, China
[3]Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana
[4]School of Information, The University of Texas at Austin, Austin, Texas
[5]Qiqihar Institute of Engineering, Qiqihar, Heilongjiang, China
[6]School of Information Engineering, Nanjing University of Finance and Economics, Nanjing, Jiangsu, China

Correspondence
Danda Ma, School of Information Engineering, Nanjing University of Finance and Economics, Nanjing, Jiangsu, China.
Email: jane.ma1030@gmail.com

Funding information
Chinese scholarship Council Project, Grant/Award Number: Student ID: 201606840093

Abstract
Collaborations are pervasive in current science. Collaborations have been studied and encouraged in many disciplines. However, little is known about how a team really functions from the detailed division of labor within. In this research, we investigate the patterns of scientific collaboration and division of labor within individual scholarly articles by analyzing their co-contributorship networks. Co-contributorship networks are constructed by performing the one-mode projection of the author–task bipartite networks obtained from 138,787 articles published in *PLoS* journals. Given an article, we define 3 types of contributors: Specialists, Team-players, and Versatiles. Specialists are those who contribute to all their tasks alone; team-players are those who contribute to every task with other collaborators; and versatiles are those who do both. We find that team-players are the majority and they tend to contribute to the 5 most common tasks as expected, such as "data analysis" and "performing experiments." The specialists and versatiles are more prevalent than expected by our designed 2 null models. Versatiles tend to be senior authors associated with funding and supervision. Specialists are associated with 2 contrasting roles: the supervising role as team leaders or marginal and specialized contributors.

1 | INTRODUCTION

In science, many solitary individuals' efforts are appreciated and emphasized. For example, people often link some individuals' names with great findings, such as Sigmund Freud with the Interpretation of Dreams, Albert Einstein with the Theory of Relativity, and John von Neumann with the Theory of Games and Economic Behavior. However, more scientific and industrial progress that has made history came from powerful collaborations. In recent decades, more and more research has been conducted by groups of scholars. For example, thanks to the joint efforts of Watson, Crick, Franklin, and Wilkins, the double-helix structure of DNA was discovered, which is fundamental to modern biotechnology (Science History Institute, 2017). Later, since 1990, 20 institutions from six countries participated in the great exploration of sequence and map of all human genes, known as the Human Genome Project. After more than 20 years of hard work, scientists were able to present nature's complete genetic blueprint of a human being; such findings greatly contribute to treat, cure, and prevent various human diseases (National Human Genome Research Institute, 2019). With the increasing complexity of problems to solve, such as designing a new functional protein or developing self-driving cars, collaboration is necessary. When checking recent leading studies, we find that many of them have a long list of contributors or acknowledgments, which reveal the intensity of collaborations. Collaboration can bring many advantages; for example, it can decrease the cost (Katz & Martin, 1997), bring in more expertise, and thus boost efficiency (Goffman & Warren, 1980), and increase scientific popularity, visibility, and recognition (O'Connor, 1970; Price & Beaver, 1966). Collaborations make the impossible possible. Many believe in the power of scientific collaborations and have spent efforts to find collaborators and work in teams (Fox & Faver, 1984).

The increasing demand for scientific collaboration has attracted numerous scholars to study the mechanism of collaboration from different perspectives, such as bibliometrics (e.g., Ding, Foo, & Chowdhury, 1998; Glänzel, 2002), social network analysis (e.g., Barabâsi et al., 2002; Newman, 2004; Zhang, Bu, Ding, & Xu, 2018), and qualitative approaches (e.g., Birnholtz, 2006; Hara, Solomon, Kim, & Sonnenwald, 2003; Lee & Bozeman, 2005;). Despite some differences, in bibliometrics most of the research uses coauthorship to measure scientific collaborations (Milojević, 2010). Studies using coauthorship usually assume that each collaborator shares equal contributions to their scientific work and based on that they build coauthor networks to study scientific collaboration (e.g., Birnholtz, 2006; Chompalov, Genuth, & Shrum, 2002; Newman, 2004). However, little is known about how each collaborating individual works in a team. Do they still collaboratively complete the whole procedure of work? Or do they divide the labor, and thus each only accomplishes certain tasks within a team and then the final goal is achieved by assembling all these tasks?

Early in the about 4th century BC, Plato stated the importance of the *division of labor* for the emergence of cities in his *Republic*; Xenophon also noticed the existence of specialization and mentioned that the division of labor enhances productivity in his *Cyropaedia*. Centuries ago, Smith (1776) discovered that *division of labor*, a proper division and combination of different operations in manufacturing, improves the efficiency of production; it further impacts our whole modern society as it shapes how people are interacting with each other to achieve goals (e.g., Durkheim, 1933; Earley, 1993; Ezzamel & Willmott, 1998). For this concern, in a team of various forms, how the tasks are divided and performed among the members determines its performance (Delfgaauw, Dur, & Souverijn, in press). Thus, there is a great need to examine the mechanisms of teamwork via investigating the division of labor in teamwork (or collaboration). In scientific collaborations, faster and greater scientific innovations are always encouraged. Therefore, it is of greater value to know how to achieve a successful scientific collaboration by a proper division of labor—which tasks each team member should take on and what kind of collaborations enables collaborators to better achieve their scientific goals (Hara et al., 2003; Ilgen, Hollenbeck, Johnson, & Jundt, 2005; Leahey & Reikowsky, 2008; Melin, 2000).

Currently, only a few have examined the scientific collaboration at the level of task assignments (Corrêa, Silva, Costa, & Amancio, 2017; Jabbehdari & Walsh, 2017; Larivière et al., 2016; Yang, Wolfram, & Wang, 2017). However, their focus was on tasks globally, rather than from the perspective of interactions within each team. For example, Jabbehdari and Walsh (2017) estimated the likelihood of specialist authors by checking the authors' tasks via a survey of 8,864 articles. Yang et al. (2017) analyzed the relationship between authors' tasks in the contribution lists and their positions in the bylines. There is a lack of research investigating the detailed division of labor within every collaboration. Here, we comprehensively analyze how members in a team divide the labor by recognizing and examining the different roles they play in a large-scale data set. Our study helps understand scientific collaborations in depth by revealing the fundamental mechanisms of how collaborative teams function. Inspired by the approaches to using the contribution statements to study author contribution patterns (Corrêa Jr et al., 2017; Larivière et al., 2016; Sauermann & Haeussler, 2017), we analyze the networks of authors and tasks from more than 130,000 articles published in *PLoS* journals. First, we study the density of the co-contributorship networks, which reflects the degree of labor division. We then define three types of author contributions—*specialists*, *team-players*, and *versatiles*—based on the co-contributorship networks, and examine the abundance of these types of contributors. We find that team-players tend to contribute to the five most common tasks, such as "data analysis" and "performing experiments." Versatiles tend to be senior authors associated with funding and supervision. Specialists are associated with two contrasting roles: the supervising role as a team leader or a marginal and specialized contributor. These features will also facilitate further assessing the division of labor and specialization in teams in the future.

2 | RELATED WORK

2.1 | Division of Labor in Teamwork Studies

Teamwork is a complex process that involves interactions between members with different expertise and skills with a spectrum of degrees of division of labor. Ilgen et al.

(2005) argued that the classic IPO (input-process-output) model is insufficient to describe the process of teamwork, due to its complexity. LePine et al.'s (2008) meta-analysis found that teamwork generally includes three general processes: mission analysis, action process, and interpersonal process; each of them includes several subprocesses. In the action process, Earley (1993) observed that the psychological statures of team members can affect their diverse collaboration patterns with others, individually or collectively. Studies also extend to classifying task types (e.g., Salas, Sims, & Burke, 2005) and team roles (e.g., Belbin, 2012). For instance, Belbin found that a team full of "Apollos" (i.e., geniuses) usually exhibits terrible performance, and that role allocation is necessary for successful teamwork.

Smith (1776) argued that division of labor is a strong impetus for increased productivity and specializations. For example, factory workers can be distributed to specific tasks in the pipeline, so that they can be more concentrated on fine-grained tasks and improve their skills (Leroy, 2009). The degree of division of labor was believed to be limited only by the number of laborers in the market (Stigler, 1951). Meanwhile, if the tasks are complex and interdependent, the coordination cost can be a significant limiting factor on specialization (Becker & Murphy, 1992). Therefore, the extent of division of labor may be largely affected by the nature of the tasks.

2.2 | Scientific Collaboration and Division of Labor

Scientific collaboration as a particular form of teamwork mainly focusing on scientific activities with high intelligence and innovation increasingly prevails in academia (Fox & Faver, 1984; Guimerà, Uzzi, Spiro, & Amaral, 2005; Katz & Martin, 1997; Larivière, Gingras, Sugimoto, & Tsou, 2015; Wuchty, Jones, & Uzzi, 2007). In this form of teamwork, division of labor is commonly suggested by some studies (Birnholtz, 2006; Fox & Faver, 1984; Kraut, Galegher, & Egido, 1987; Leahey & Reikowsky, 2008). For example, Melin (2000) classified scientific teams into two categories: one where everyone in the team is given a clear task assignment and the other where everyone works together. The two types are defined in a similar way by Hara et al. (2003) as "complementary" and "integrative" teams. Chompalov et al. (2002) classified teams into four categories based on their topological features: bureaucratic, leaderless, nonspecialized, and participatory teams.

Current studies investigate scientific collaboration via coauthorship network analysis (Ahuja, 2000; Newman, 2004; Xie, Ouyang, Li, Dong, & Yi, 2018; Yan & Ding, 2009) or using case studies (Amabile et al., 2001) and interviews (Birnholtz, 2006; Chompalov et al., 2002; Chung, Kwon, & Lee, 2016; Fox & Faver, 1984). These studies reveal several important features in collaborations, such as homophily (Zhang et al., 2018), transitivity (Newman, 2004), and preferential attachment (Milojević, 2010). They also suggest that collaboration improves productivity in science (Lee & Bozeman, 2005) and collaborative research tends to attract more citations (Larivière et al., 2015). However, such coauthorship studies usually overlooked the division of labor in scientific collaboration at large; and some of them only relied on a limited number of cases. Only a few studies started investigating the tasks conducted by the members of a team. But there is still a lack of research investigating the roles scientists have played in every collaboration. These drive us to use author contribution statements embedded in the full text of scientific articles provided by authors to investigate how scientific teams design their tasks and distribute them to collaborators, which is the process of division of labor. Thus, we can investigate the scientific collaboration between coauthors at the task level and reveal different roles played by these authors.

2.3 | Contribution Statement for Scientific Collaboration Studies

Although the author contribution patterns in scholarly articles have been of interest in scientometrics (e.g., Giles & Councill, 2004; Laudel, 2002), it was the wide adoption of the contribution disclosure policies that enabled large-scale data-driven studies (Allen, Scott, Brand, Hlava, & Altman, 2014; Brand, Allen, Altman, Hlava, & Scott, 2015; Larivière et al., 2016). For example, Larivière et al. (2016) examined the forms of division of labor across disciplines, the relationship between contribution types (i.e., writing the article, performing the experiments, conceiving ideas, analyzing data, and contributing tools) and authors seniority, such as academic ages and that between types of tasks and byline positions. They found that authors contribute to their studies unevenly across disciplines; that most authors are identified to contribute to writings; and that those who write the articles usually design the studies and those providing materials usually do not perform an experiment, and vice versa. They also found that senior authors usually do fewer tasks such as conducting experiments than junior ones, but do more tasks such as writing articles and contributing tools and materials. First and last authors usually contribute more tasks than middle ones to their studies. Corrêa Jr et al. (2017) placed more emphasis on the relationship between authors' rank positions and their corresponding contributions. They collected author contribution statements in *PLoS One*, identified five common tasks, and built a bipartite graph for each article, where authors and the five tasks are the two groups of nodes and an edge between author and task means the author performed the task, treating tasks as equal contributions. Using the average number of tasks authors performed across articles, they found that usually the first and the last authors contribute more to their articles than middle authors, which echoes the findings by Larivière et al. (2016). They further identified three general patterns of author contribution with their byline position: the contribution increases with authors' ranks, the contribution decreases with authors' rank, and the contribution decreases then increases with the author's ranks. Sauermann and Haeussler (2017) presented two studies: the first investigated how informative the byline position of an author is about the type and broadness of the author's contribution using more than 12,000 *PLoS One* articles; the second reported how author contribution statements are used and scholars' several concerns on authorship and author contribution statement after surveying nearly 6,000 corresponding authors from *PLoS One* and PNAS (*Proceedings of the National Academy of Sciences of the United States of America*). The two data sources suggest no significant differences. They also found similar observations that the first and the last authors contribute more than the middle authors to their articles (Corrêa Jr et al., 2017; Larivière et al., 2016). In addition, they also observed that corresponding authors are more likely to be the last authors. First authors usually tend to make more contributions than other authors. When the team gets larger, authors tend to perform fewer tasks, suggesting a stronger degree of division of labor. The top 10% most cited articles maintain similar results from the models generated from the full data set, suggesting the reliability of the author contribution statements from *PLoS One* articles.

To sum up, as the division of labor has been an important driving force in modern society (e.g., Durkheim, 1933; Earley, 1993; Ezzamel & Willmott, 1998), there has been much interest in studying the division of labor or roles in teams, particularly in scientific collaboration teams. The author contribution statements can serve as a good proxy to concretely measure the role allocation and division of labor in individual scientific collaborations. Given the complex nature of scholarly work, it is of great value to ask how a team can achieve a successful scientific collaboration, how the division of labor occurs in scientific collaboration, and what the patterns of role and labor distribution are (Hara et al., 2003; Ilgen et al., 2005; Leahey & Reikowsky, 2008; Melin, 2000).

3 | METHODOLOGY

This section explains our data set and approach to constructing and analyzing the contribution network. First, we collected 138,787 full-text articles from *PLoS*, from which we extracted and parsed the author contribution statements. From each statement, we extracted author–task pairs, which we assembled to construct an author–task contribution network for each article. The one-mode projection of this network produces a co-contributorship network, from which we define the three types of collaborators. Finally, we further investigated the tasks they partook in using content analysis.

3.1 | Data

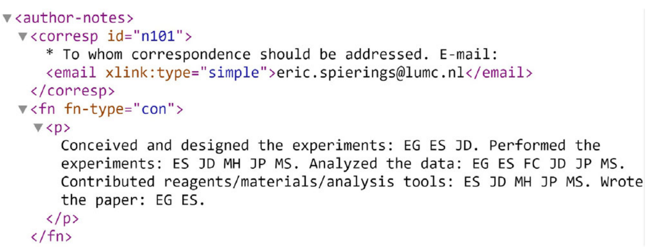
3.1.1 | Full-text data source

PLoS is one of the largest open-access journal article publishers in the world. Under its authorship policy[1] (which accords with CRediT Taxonomy[2] since 2016), authors are required to state their agreed contributions in their articles. To collect author contribution statements, nearly 170,000 full-text articles published in *PLoS* between 2006 and 2015 were harvested in XML format.

In each XML file, the author contribution statement is either embedded in the tag of "*<fn fn-type = 'con'>*" (see the sample article in Figure 1) or in the acknowledgement part (a few of them).

3.1.2 | Author–task pairs

We used the XML package in Python 2.7 to extract the contribution statements. Then we used regular expressions to extract author–task pairs from each statement of every article. Table 1 shows the parsed data from one sample article. We did not separate commonly grouped tasks into subtasks. For instance, we consider



```
▼<author-notes>
  ▼<corresp id="n101">
     * To whom correspondence should be addressed. E-mail:
     <email xlink:type="simple">eric.spierings@lumc.nl</email>
   </corresp>
  ▼<fn fn-type="con">
    ▼<p>
       Conceived and designed the experiments: EG ES JD. Performed the
       experiments: ES JD MH JP MS. Analyzed the data: EG ES FC JD JP MS.
       Contributed reagents/materials/analysis tools: ES JD MH JP MS. Wrote
       the paper: EG ES.
     </p>
   </fn>
```

FIGURE 1 An example author contribution statement of a sample article in XML format [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Author–task pairs of a sample article

Id	Authors	Task
1	EG; ES; JD	Conceived and designed the experiments
2	ES; JD; MH; JP; MS	Performed the experiments
3	EG; ES; FC; JD; JP; MS	Analyzed the data
4	ES; JD; MH; JP; MS	Contributed reagents/materials/tools
5	EG; ES	Wrote the article

TABLE 2 Journal distribution of our collected author contributions

Journal	# of articles	Ratio (%)
PLoS One	12,422	89.5%
PLoS genetics	3,919	2.8%
PLoS pathogens	3,445	2.5%
PLoS computational biology	3,067	2.2%
PLoS neglected tropical diseases	2,783	2.0%
PLoS biology	921	0.7%
PLoS medicine	432	0.3%

"Contributed reagents/materials/tools" or "conceived and designed the experiments" as single tasks of their own.

The final collected 138,787 articles belong to seven journals in *PLoS*, as shown in Table 2. The table suggests that 90% of articles are from *PLoS One*, and the other 10% belong to the other six journals by *PLoS*. The distribution of those articles by year in our data set is presented in Figure 2, suggesting most of the articles were published before the middle of the year of 2016, when the CRediT Taxonomy was launched for regulating author contribution statements.

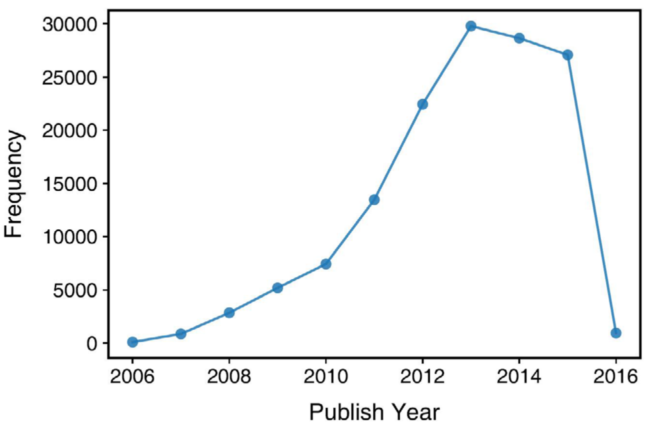


FIGURE 2 Yearly distribution of articles in our data set [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | Co-contributorship Network Construction

3.2.1 | Definitions

Figure 3 illustrates different types of collaboration patterns that one can observe from co-contributorship networks. In Figure 3a, every author worked collectively on each task, forming a complete graph. Under this scenario, the division of labor does not occur, as everyone works on all tasks collectively. By contrast, in Figure 3c, every author worked on his/her tasks independently, thus having a strong division of labor. In our data set, we expect to see the whole spectrum from no-division to complete division, while most collaborations would occur somewhere in the middle (e.g., Chompalov et al., 2002; Fox & Faver, 1984; Heffner, 1979).

Building on this intuition, we formally consider a weighted undirected co-contributorship network for each article, which can be obtained by performing a special one-mode projection to the author–task network. This process is different from the standard one-mode projection because we also create self-edges if a task is performed by only one person. In the co-contributorship network, each node represents an author (collaborator). An edge between authors means that there is at least one task where two authors collaborated. The weight of each edge represents the number of tasks co-performed by the two authors. If a task is performed by more than two authors, every possible pair of authors will have an edge between them. If a task is performed by a single person, the node (collaborator) will have a self-loop, and its weight is decided by the number of tasks that the author performed independently. As an example, Figure 3d demonstrates a co-contributorship network between four authors in one article. The weight of the edge (C2, C3) is three, which means authors C2 and C3 worked together on three different tasks. The weight for the self-edge of C1 is two, indicating that C1 independently worked on two different tasks alone.

3.2.2 | Types of collaborators

Based on its connectivity patterns, each node is classified into one of the three roles: team-players, specialists, and versatiles, as shown in Figure 3e. Team-players are those who do not have any self-edges; they performed all their tasks with someone else. Specialists are those who have only self-edge(s) (e.g., C1); they are those who finish their tasks on their own. Versatiles are those who have both self-edges and normal edges (e.g., C2).

3.2.3 | Null models

To estimate the expected prevalence of each type of collaborators, we adopt two null models to the author–task contribution networks: the configuration model (Molloy & Reed, 1995) and the Erdős–Rényi random graph model (Erdös & Rényi, 1959). In the configuration model (CFM), the degrees of nodes are fixed, while the actual connections are randomized. In creating the networks, we reject the cases with multi-edges. Finally, we project this author–task bipartite graph to a co-contributorship network (see Figure 4a). For the Erdős–Rényi random graph model (ERM), we fix the number of edges in the author–task graph and randomize the connections without conserving the degree sequences. To make the random graph realistic, we enforce the

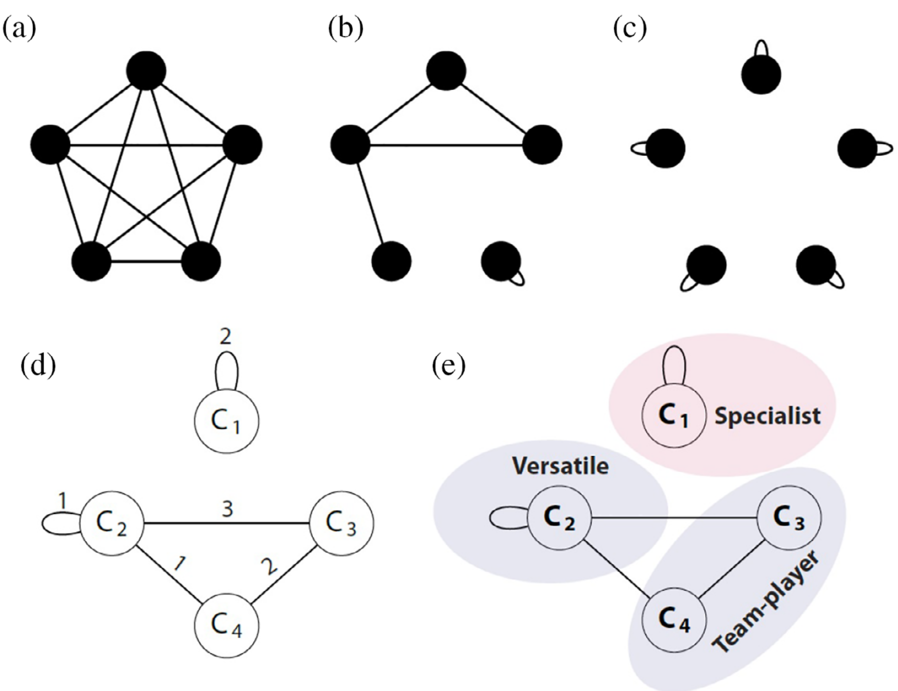


FIGURE 3 Modes of division of labor in teams (**a–c**); co-contributorship network (**d**); and types of collaborators (**e**) [Color figure can be viewed at wileyonlinelibrary.com]

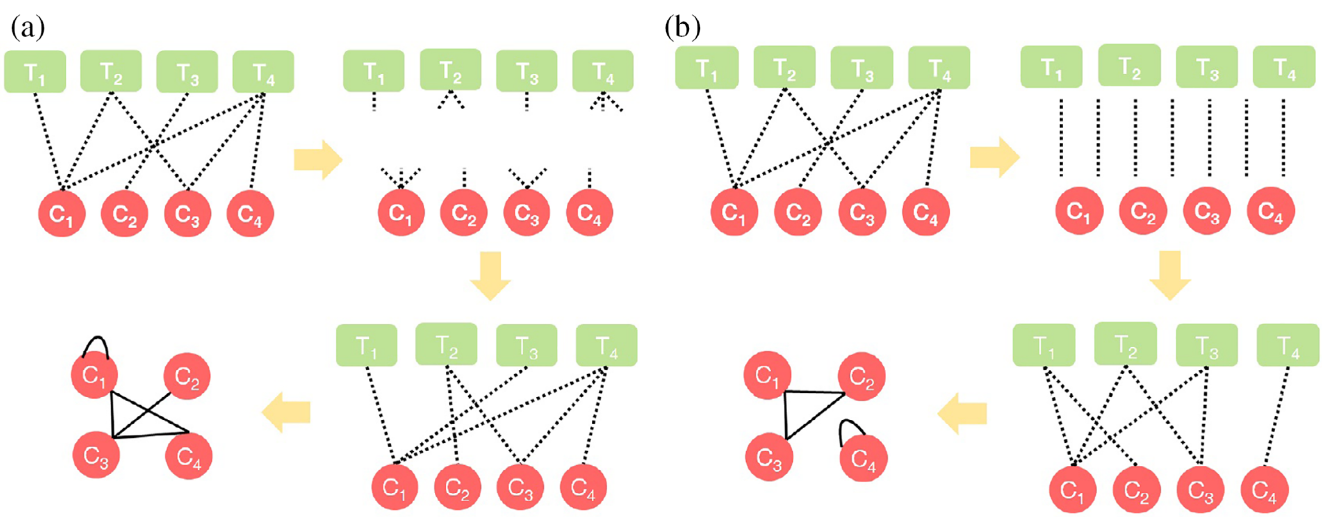


FIGURE 4 Producing null models using CFM and ERM. (**a**) Configuration model (CFM), where the degree sequences on both sides are preserved. (**b**) Erdős–Rényi random graph model (ERM), where only the total number of edges is preserved with every node having at least one connection [Color figure can be viewed at wileyonlinelibrary.com]

connectivity of the network—each author node and each task node should have at least one edge. After obtaining an initial random graph, we perform a rejection sampling to obtain a graph where every node has at least one connection (see Figure 4b). By examining the differences between the actual networks and the two null models, we put our measurements in a reasonable context of random cases.

3.3 | Research Hypotheses

Using the networks we built above, we sought to answer three questions concerning the division of labor within teams.

RQ1 *Is division of labor common in scientific collaborations?*

To answer this question, we examined the density of each bipartite graph we built and compared it to the expectation from their corresponding ER random graphs, which maintains the number of edges between authors and tasks. Ideally, if the division of labor is not necessary for scientific collaboration, the graph density distribution of all the networks we built will follow a binomial distribution, where the chance to connect an author and a task in an author–task bipartite network is equal. So our first null hypothesis for this question will be:

H01 *There are no differences in the graph density distribution between the real-world author–task bipartite networks and the randomised ones.*

RQ2 *Concerning the three types of collaborators, are they more common than one another in scientific collaboration?*

To answer this question, we examined the distribution of the three types of collaborators in their scientific collaborations at an article level from three perspectives: the existence of the collaborators, the ratios of them in all of the publications, and the ratios of them in the publications with nonteam-players. We also designed the ER and the CRF models for each co-collaboratorship network to remove random factors from the observations. So our null hypotheses for this question will be:

H2 *The three types of collaborators are equally common in scientific collaborations.*

H3 *The ratios of the three types of collaborators in all publications are equal to each other.*

H4 *The ratios of the three types of collaborators in the publications with nonteam-players are equal.*

RQ3 *Do the three types of collaborators perform different tasks in their collaborations against each other?*

To answer this question, we examined the distribution of the tasks that the three types of authors performed in their scientific collaborations at an article level in two parts: the five common tasks, and the rest less frequent tasks in all publications. The ER and the CRF models served to remove random factors from the observations. So our null hypotheses for this question will be:

H5 *The three types of collaborators contribute equally to the five common tasks in their scientific collaborations.*

H6 *The three types of collaborators contribute equally to the less frequent tasks in their scientific collaborations.*

Following the three questions above with six null hypotheses, we used the author co-contributorship networks built from each article's author contribution statement to address these questions and hypotheses.

4 | RESULTS AND DISCUSSION

4.1 | Overview

More than 90% of the articles in our data set were written by at least two authors, agreeing with the previous observations that collaborative studies are dominating (Guimerà et al., 2005; Wuchty et al., 2007). 87% of articles were written by teams of no more than 10 members; and 99% of teams had no more than 20 authors, including 1,370 single-authored articles (8.2%), shown in Figure 5a. In the following analyses, we focus on the articles with fewer than 20 authors in our data set because they occupy the vast majority of the data set and it is easier to implement the null models for the articles with fewer authors.

To depict the division of labor in scientific collaboration, we first calculate the normalized graph density of each author–task bipartite graph compared with one null model generated by Erdős–Rényi random bipartite graph (detailed below). The normalized graph density is calculated using Equation (1):

$$NGD = \frac{k\text{-max}(N_t, N_a)}{N_t \times N_a\text{-max}(N_t, N_a)} \tag{1}$$

where k represents the number of edges in the graph, N_t number of tasks, and N_a the number of authors. So

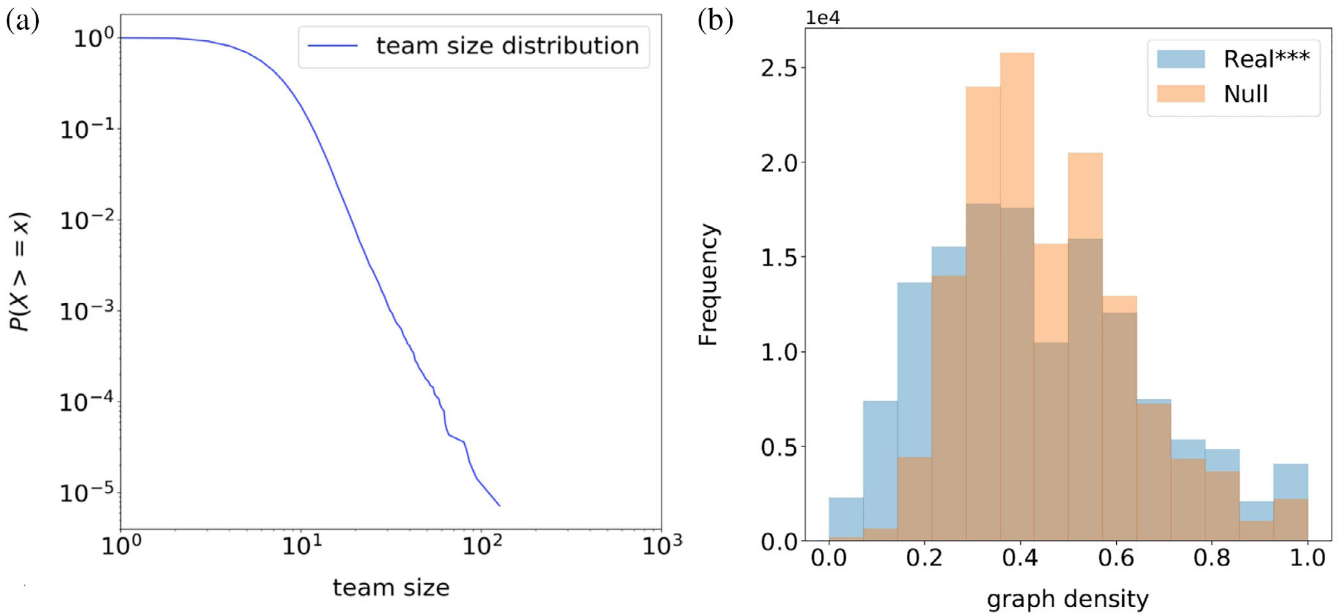


FIGURE 5 Author distribution by team size (**a**); and normalized graph density distribution in author–task bipartite graph (**b**) (*denotes the p-value from Kolmogorov–Smirnov test against the null model, *p-value <.05, **p-value < .01, ***p-value < .001) [Color figure can be viewed at wileyonlinelibrary.com]

$N_t \times N_a$ denotes the maximum number of edges in an author–task bipartite graph and $\max(N_t, N_a)$ represents the minimum when all nodes should be connected.[3]

To generate our null model here, another Erdős–Rényi random bipartite graph was adopted, using $G(N_t, N_a, p^j\langle N_a\rangle)$ where $p^j\langle N_a\rangle$ is the probability for an author to perform a task in article j which contains N_a collaborators (Batagelj & Brandes, 2005), estimated by using Equation (2):

$$p^j\langle N_a\rangle = \frac{N_e^j/N_a}{\bar{N}_t} \tag{2}$$

In Equation (2), N_e^j is the number of edges in the author–task bipartite graph of article j; $\bar{N}_t$ is the mean number of tasks in all articles with N_a collaborators; and N_e^j/N_a the average number of tasks per collaborator performed in N_a-author article j. Then we use Equation (1) to calculate the normalized graph density for these random graphs.

Figure 5b shows that the author–task bipartite graphs in our data set present larger variance in the degree of labor division, compared with the null model that assumes a homogeneous contribution from authors. By examining two ends of the x-axis, it is found that both a strong division of labor and no division of labor are more probable than expected by the homogeneous null model. This might suggest that scientific teams tend to employ a wider variety of collaboration strategy, although our results may be explained by the heterogeneous author degree distribution (i.e., large variation in the number of tasks one performs).

Figure 6 shows the graph density of groups with a different team size. A clear trend can be observed that the graph density distribution of the real-world author–task bipartite graphs is more and more divergent from the density expected by the null model when the team size grows. Specifically, the author–task graphs in real collaborations tend to be sparser than expected in the null model, which might suggest a stronger degree of labor division in larger teams.

4.2 | Quantifying Types of Collaborators

In this section we examine the prevalence of the three types of collaborators. First of all, we examine how many articles involve these three types of collaborators. We calculate the ratio of the articles containing *collaborator type* c_i, given *team size* k, $PR^k_{C_i}$, using Equation (3) as follows:

$$PR^k_{C_i} = \frac{N^k_{c_i}}{N^k}, c_i \in \{specialist, versatiles, team-player\} \tag{3}$$

where N^k is the number of articles with k authors, and $N^k_{c_i}$ is the number of articles with k authors that contain *collaborator type* c_i.

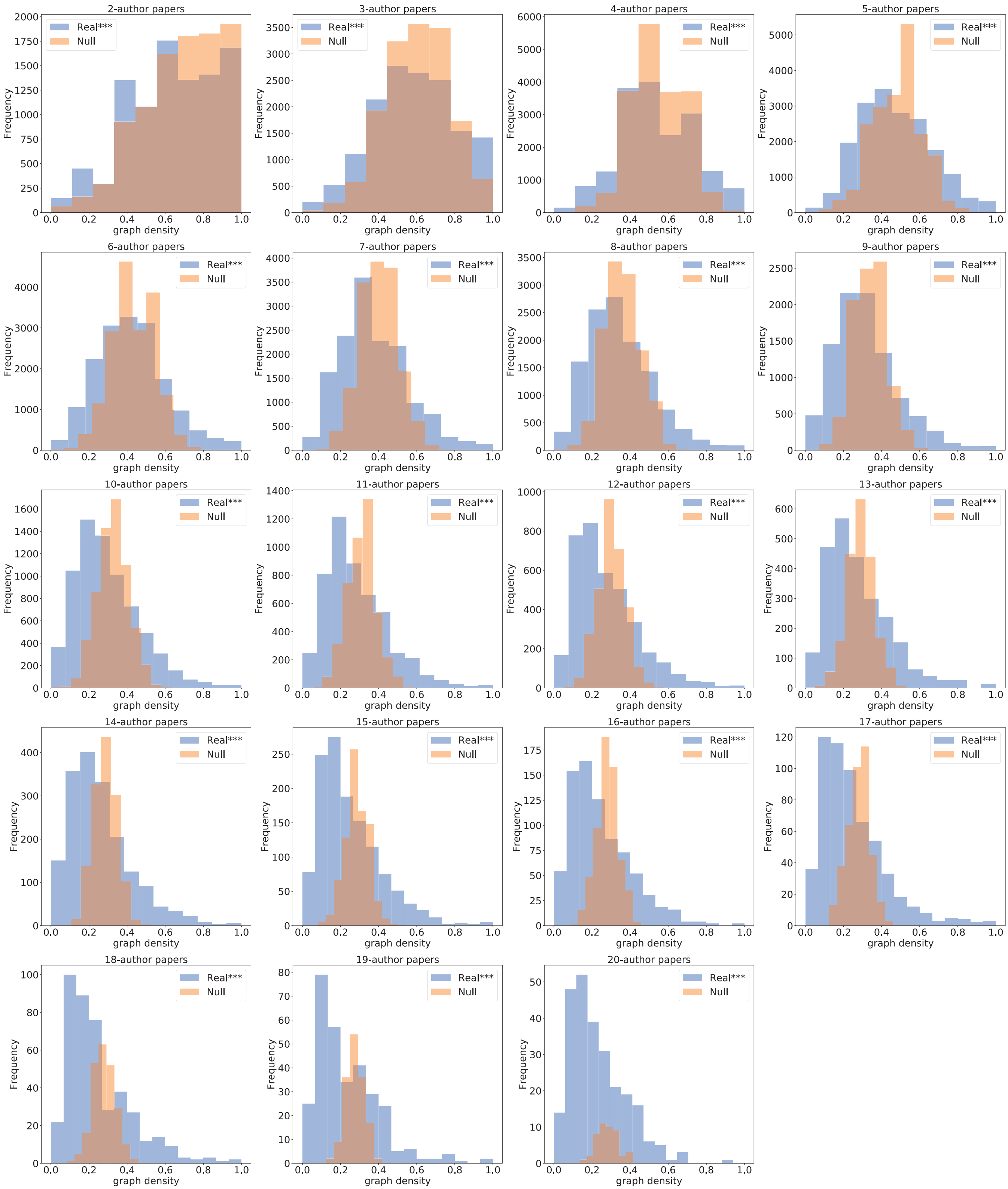


FIGURE 6 Graph density against null model by team size (*denotes the *p*-value from Kolmogorov–Smirnov test against the null model, **p*-value < .05, ***p*-value < .01, ****p*-value < .001) [Color figure can be viewed at wileyonlinelibrary.com]

Figure 7 shows that most articles have team-players, and that is expected by both null models (in A and B). There is a slight increase in the number of articles with specialists as team size increases. The articles with versatiles become less common as the size of teams increases, and the final ratio of such articles stabilizes around 25%.

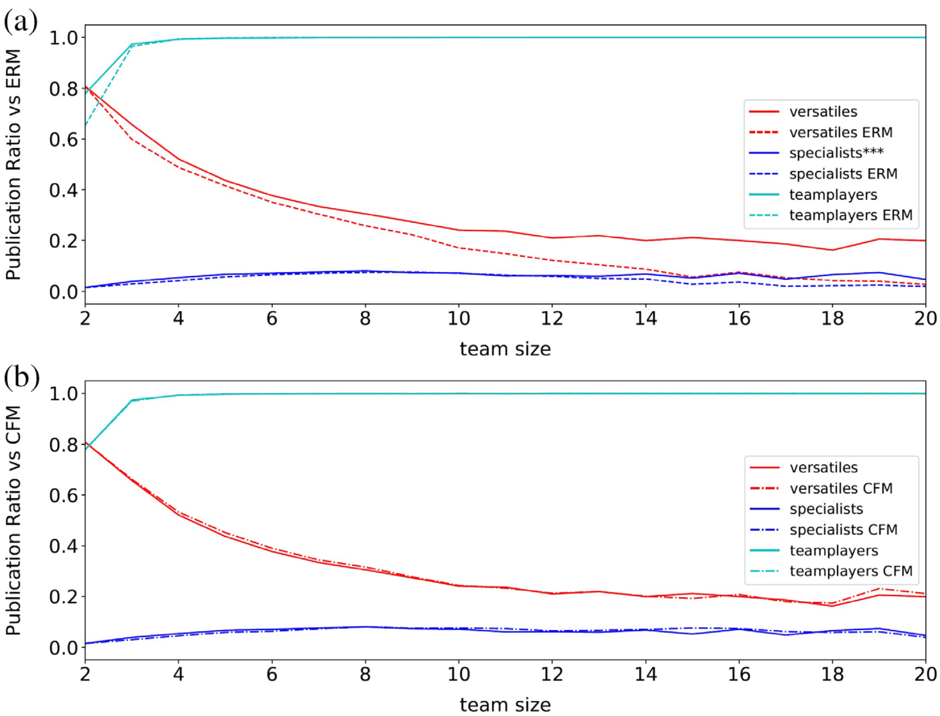


FIGURE 7 Frequency distribution of articles containing given collaborator type by team size: (**a**) against ERM null model; and (**b**) against CFM null model (*denotes the *p*-value from K-S test against the null model, *p-value < .05, **p-value < .01, ***p-value <.001) [Color figure can be viewed at wileyonlinelibrary.com]

When compared with the ERM null model, which shuffles author and task nodes in author–task bipartite graphs, nonteam-players are more common in real-world collaborations than expected. More articles involve versatiles in real scientific collaborations. Specialists, instead of disappearing in larger teams, as ERM suggests, keep playing a role in teams whose sizes vary from 2 to 20. This might suggest that nonteam-players are associated with special and prevalent types of contributions in scientific collaborations.

Our results also suggest that the actual prevalence of each author type closely matches the expectation from the CFM null model, which shuffles authors' specific contributions in the bipartite graphs. This indicates that the degree sequence—how many tasks are performed by each author—accurately reproduces the co-contribution patterns that are observed.

To observe three types of collaborators' existence in scientific collaborations, we calculate the average number of each type of collaborators $C_i$in teams by team size k, using Equation (4) as follows:

$$AC_i^k = \frac{TC_i^k}{N^k}, c_i \in \{specialist, versatile, team-player\}, 2 \le k \le 20 \tag{4}$$

where N^k is the number of teams with team size k, and TC_i^k is the total number of *collaborator i* in k-authored publications. The results in Figure 7A suggest that, on average, each article contains around 0.075 specialists, 0.35 or more versatiles, and the rest are team-players when the team size is greater than five. Specifically, when the team size is smaller than eight, the number of specialists increases along with the increase of team size and peaks at 0.1, which is well captured by both our null models. When team size continues growing, the real number of specialists fluctuates around the top, whereas that expected by ERM starts to diminish (shown in Figure 8a, left). Versatiles are more prevalent than specialists in scientific collaborations, especially among smaller teams. When the team grows larger, the average number of versatiles continues decreasing to 0.35 per article, then maintains stable. Despite the decreasing average number of versatiles in teams, it is still more than expected by ERM, in which the average number of versatiles keeps declining when the team size is larger than 10. Team-players dominate the participation in scientific collaborations, which is well captured by our two null models. The ERM null model only keeps the number of task assignments, while the CFM model sets some rules of labor division, since it restricts how many tasks one author would participate in and how many participants are involved for each individual task. By comparing the figures of the real situation in Figure 8a with the two null models, we could see that there exists a strong division of labor and that there are still more specialists and versatiles than we would expect from the random cases among the collaborators. A downtrend of the number of versatiles along with the increase of team size is understandable, since the total number of tasks will not have an unlimited growth, so some authors may collaborate more with others when there are more team members. However, the slight

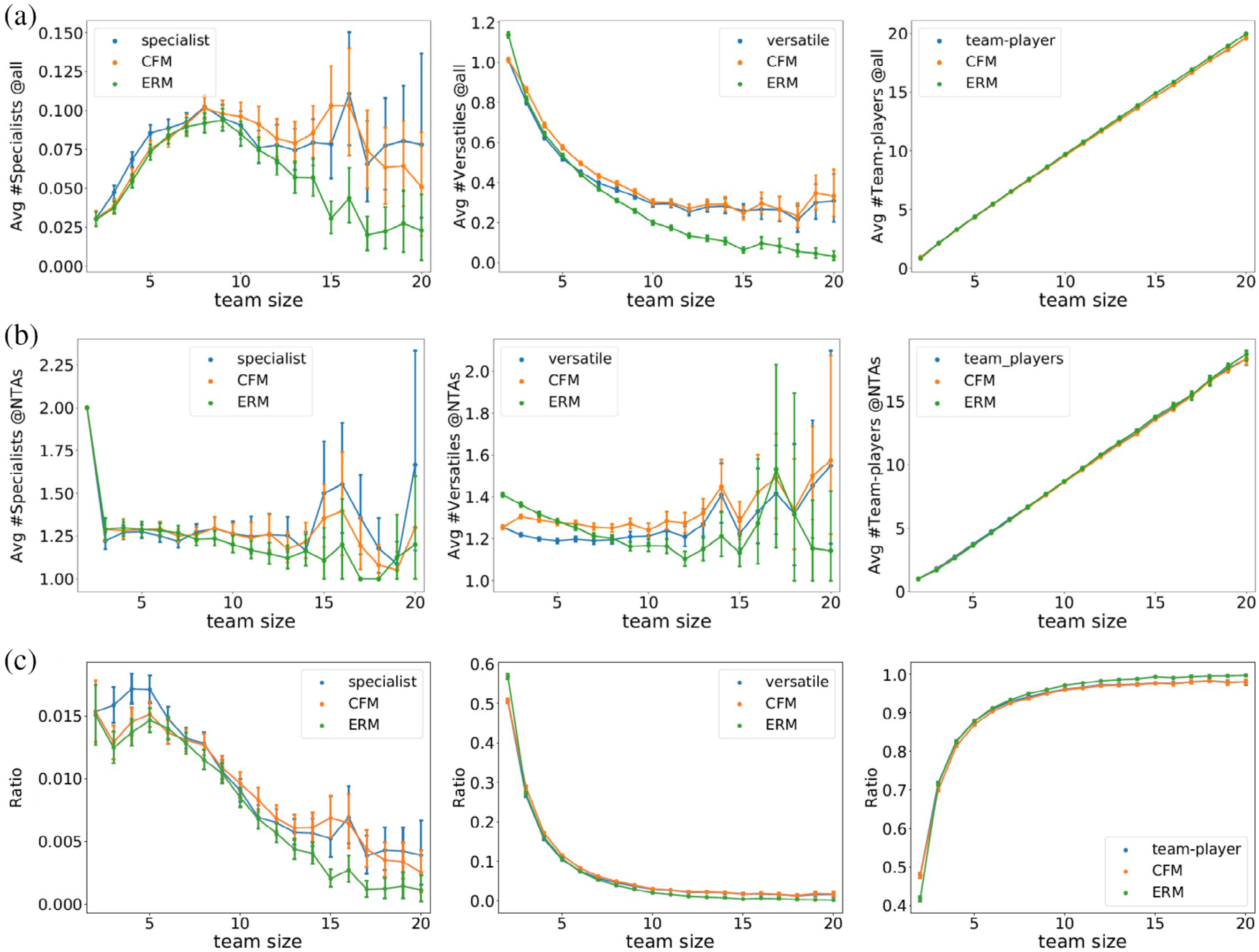


FIGURE 8 (**a**) Types of collaborators in publications; (**b**) Team structure in NTAs (nonteam-player-involved articles); and (**c**) Population ratio of types of collaborators. From left to right in each subplot are specialists, versatiles, and team-players; error bars in plots represent 95% confidence intervals generated via 10,000 bootstrapping iterations [Color figure can be viewed at wileyonlinelibrary.com]

increase or unchanged number of specialists demonstrates that there always exist some tasks that should be completed individually; the existence of specialists is important even in the environment of heavy collaboration. The distinction between the figures of nonteam-players for the real-world collaborations and ERM indicates that the existence of nonteam-players is not because of small teams or limited labors in scientific collaborations but for particular purposes left for us to uncover. We are more interested in understanding the structure of scientific teams when they involve heterogeneous collaborators. Thus, we exclude all the articles that were collaborated by only team-players and plot the average number of three different collaborator types among the remaining ones in Figure 8b. In general, the team structure is quite stable with nonteam-players among all different team sizes: 1.25 specialists on average, 1.3 versatiles on average, and with the rest team-players. More than often, a team includes one or two nonteam-players to perform their research. Despite that, when teams grow larger than 15 participants, more nonteam-players, specialists, or versatiles, could contribute to the teams (suggested by the error bars). Team-players, similarly, still dominate a team. The null models, however, do not show great disparity from the real collaborations in the team structures. This indicates when teams include nonteam-players, there is no big variance among teams, especially for smaller teams whose sizes are less than 10.

We continue our focus on the overall population of the three types of collaborators among all the publications. We modify Equation (4) and calculate RC_i^k, the ratio of *collaborator C_i* given by *team size (k)*, using Equation (5) as follows:

$$RC_i^k = \frac{TC_i^k}{k \times N^k}, c_i \in \{specialist, versatile, team-player\}, 2 \le k \le 20 \quad (5)$$

where N^k is the number of teams with team size k, and TC_i^k is the total number of *collaborator* i in k-authored publications. The results of Equation (4) are the results from Equation (3) normalized by team size accordingly (shown in Figure 8c). Since the authors in our whole data set are not disambiguated, the population character of these collaborators reflects how frequently a certain role (as three types of collaborators) has been played in scientific collaborations.

Figure 8c demonstrates that nonteam-players are the minority in scientific collaborations, as suggested above, especially for specialists. In particular, when the team size grows, the ratio of specialists among collaborators drops from 15% to 0.5%, then remains stable; the ratio of versatiles also falls from 55% to 3%, then remains stable. Team-players, on the contrary, show an opposite trend, keeping an increase from around 45% to around 95%. Both our two null models also roughly capture this trend.

To sum up, team-players are the major collaborators in scientific collaborations. Nonteam-players are the minority, but they widely exist in small teams (a size no more than five) and also exist in larger teams (a size larger than five) with a relatively small and stable ratio.

A possible reason for these observations is that more team members enable division of labor and specialization (Smith, 1776) rather than wiping out nonteam-players. Some of the tasks performed by versatiles in smaller teams can be distributed to extra team members, accounting for more team-players. Regarding specialization, some team members can focus on particular tasks when more members are added to the team. For instance, in a dyadic team between advisor and advisee, besides supervision, the advisor may also need to take up some tasks such as writing and data analysis to accelerate the research progress. When more collaborators get involved, the advisor may spare more time and only focus on the supervision of advisees and funding application. Other collaborators can share the burden of the advisor (Bray, Kerr, & Atkin, 1978) when the advisor could be a specialist and the other collaborators can function as team-players. In addition, specialized collaborators can be invited to the team to perform some special tasks as specialists (or Specialists proposed in Belbin, 2012). The benefit of this evolution—division of labor and specialization—can increase the productivity of a team. On the other hand, however, more collaborators could bring in the so-called "Ringelmann effect" (Ingham, Levinger, Graves, & Peckham, 1974) or "social loafing" (Earley, 1993), which means collaborators of a team tend to become increasingly less productive as the size of their team increases. However, this increasing tendency of specialization reaches saturation instead of excessively extending, which might be taken as the consideration of a huge coordination cost that specialization may lead to (Becker & Murphy, 1992).

Randomly assigning tasks to authors (in ERM) leads to more nonteam-players in smaller teams (≤10) and less in larger teams (≥15). By contrast, in real scientific collaborations, nonteam-players maintain a relatively stable ratio in smaller teams and also exist in larger teams. Such existence is surprising in larger teams, since adequate human resource facilitates us to perform tasks collaboratively to achieve seeming efficiency and effectiveness.

Also, it is worth noticing that versatiles tend to be more favorable than expected in scientific teams as suggested by Figure 7. On the contrary, fewer versatiles in the ERM plot may imply that team-players are more welcome when they can also work independently as versatiles. The possible reason for this can be that a moderate degree of specialization improves the efficiency of the collaboration when some tasks are performed alone and some collaboratively (Becker & Murphy, 1992).

4.3 | Understanding Collaborators' Tasks

After quantitatively describing the prevalence of the three types of collaborators in our data set, here we analyze their characteristics by examining the tasks in which they participated. We look at the most common five tasks (e.g., Corrêa Jr et al., 2017; Larivière et al., 2016) as well as the other less frequent tasks. Using the data generated by our null models used in the previous section, that is, CFM and ERM, we can also investigate the different patterns in task distributions between real collaborations and two random scenarios for different purposes. CFM controls authors' and tasks' degree sequence in an author–task network; thus, the differences from real collaborations highlight the differences in task-performing, which will suggest prevailing patterns of different types of authors in reality. ERM only controls the number of edges in the networks. The corresponding results can be used to examine whether these task patterns can be generated randomly. We extract the top 100 most frequent tasks for each type of collaborators from the three data sets (including the two generated null model data sets) and consolidate similar tasks. As a result, 52 unique tasks are obtained.

4.3.1 | Five common tasks

Figure 9 presents the five most common tasks in the author contribution statement from *PLoS*. The radar plots suggest that although the three types of authors all engage in the five most common tasks, emphases vary. For instance, specialists "contribute reagents/materials/analysis tools" much

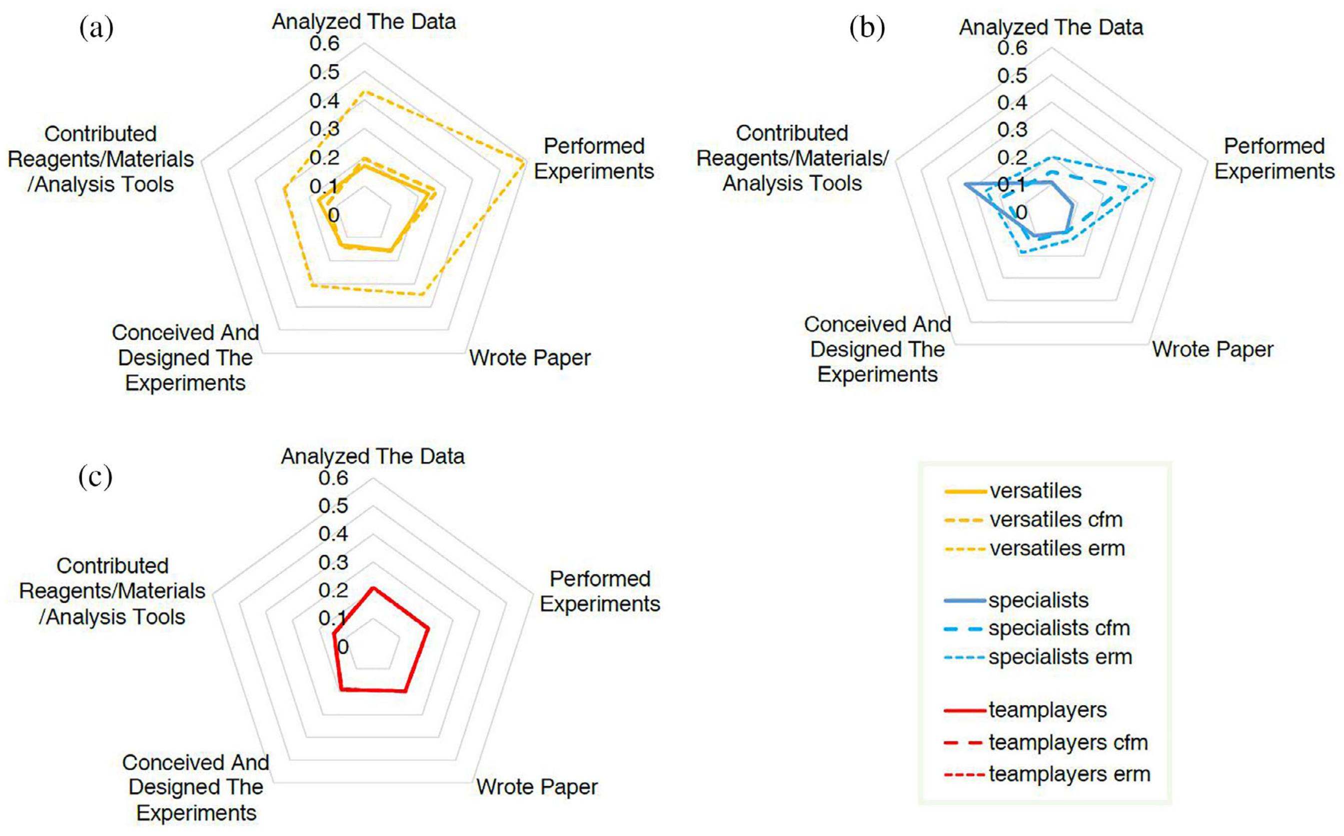


FIGURE 9 Common tasks performed by the three types of collaborators: (**a**), Specialists; (**b**), Versatiles; (**c**), Team-players [Color figure can be viewed at wileyonlinelibrary.com]

more often than expected, while "performed experiments" much less than expected (Figure 9a). This result suggests that the "reagents/materials/analysis tools" task can be more easily isolated to a single person than other tasks, and that it is rare for an author to just perform experiments and not participate in other tasks with others, indicating the central role of the task of performing experiments in scientific studies. Compared with null models, specialists make much less contribution to performing experiments. This might indicate that specialists are not usually implementers but toolmakers in a team. On the contrary, versatiles and team-players show more balanced contributions to all the five common tasks. Despite that, the null models suggest that versatiles contributed much less to the five common tasks than expected, indicating their emphasis on less common tasks. Team-players show indifferences in null models, which could be attributed to their massive population in our data set.

4.3.2 | Less frequent tasks

Figure 10 and Figure A1 (in the Appendix) show the participation patterns of the less frequent tasks. Greater disparities emerge, which have been overlooked by existing studies that focus only on a few core tasks (e.g., Corrêa Jr et al., 2017; Larivière et al., 2016). First, team-players participate in these activities much less frequently, except for "approve the article," which usually occurs at the final stage of their research. "Data interpretation" is another task that team-players do more frequently than nonteam-players. The possible reason for this is that data interpretation is interdependent with data analysis, which is a team-players' major task in Figure 9c. Similar to Figure 9c, both null models capture almost identical patterns for less frequent tasks. Second, specialists show a strong tendency to take the tasks like "review article," "revised article," and "supervised the research." This might indicate that specialists can be senior investigators in teams. Some of the following tasks, like "principal investigator" and "provided guidance," also suggest our inference. The CFM confirms that specialists contributed more to these tasks as senior authors than expected, such as "revised article" and "supervised the research." In addition, specialists also perform tasks that may not be that crucial, such as "collected data" and "collected samples." This type of task may also suggest that specialists can be mild collaborators (Hara et al.,

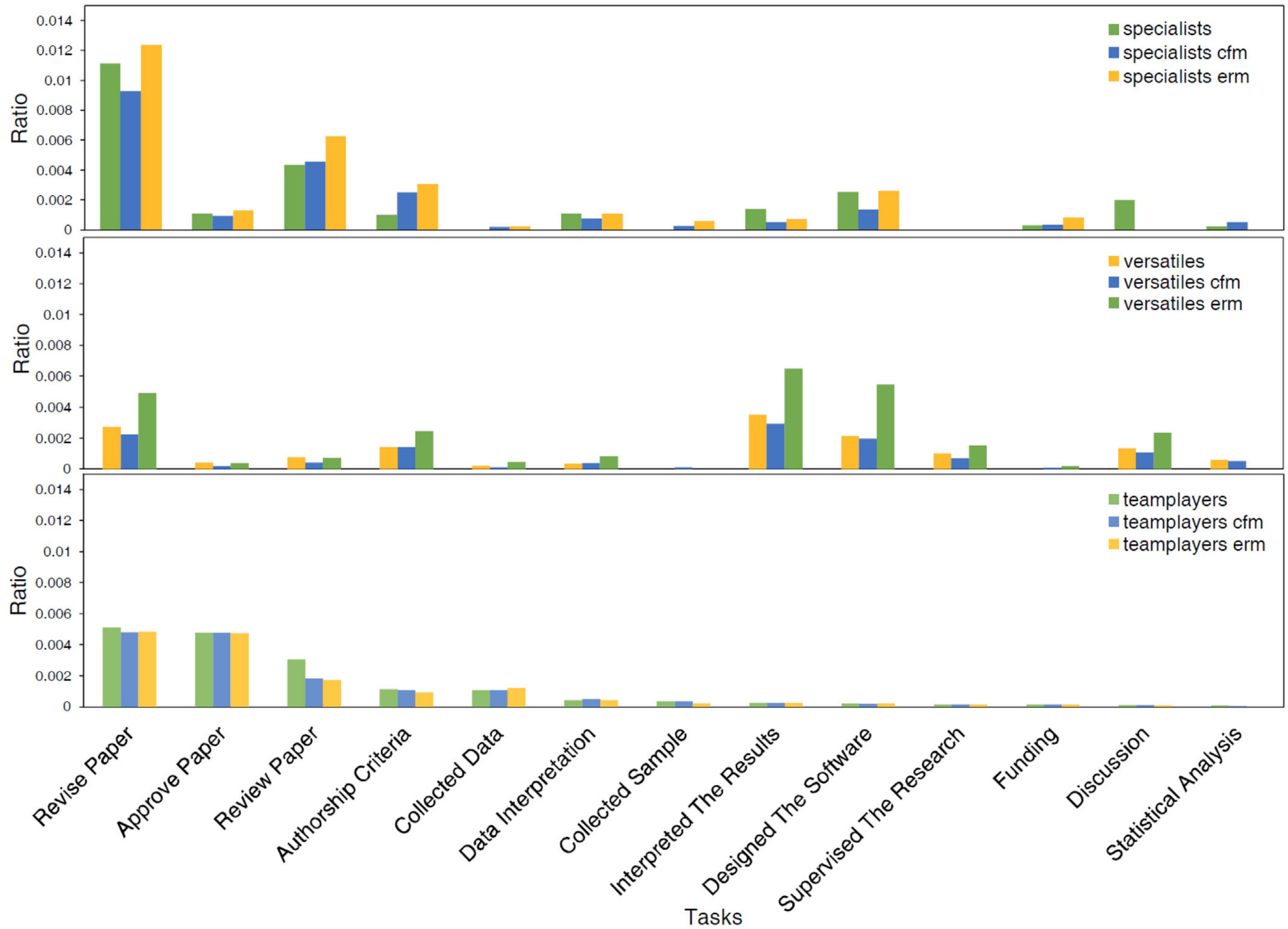


FIGURE 10 Less frequent tasks performed by the three types of collaborators (more information about less frequent tasks is presented in Figure A1) [Color figure can be viewed at wileyonlinelibrary.com]

2003). As suggested by the following tasks as well, they also take charge of "database management" and "provided technical support" (in Figure A1). CFM also confirms this tendency. Versatiles tend to partake in authority-intensive and idea-intensive tasks. For example, most of the funding-related tasks are versatiles' work. Designing software and designing models are usually versatiles' tasks. This is confirmed by our CFM. We may infer that some versatiles are either leaders of certain projects or chief authors of the studies.

5 | CONCLUSION

In this study we proposed a refined approach—author contribution network for each publication based on the author contribution statements embedded in the body of articles. It aims at better understanding scientific collaboration at the task assignment level. More than 130,000 articles were collected to perform our analyses.

The results suggested that scientific collaboration within the team could be diverse. Inspired by the concepts of division of labor and specialization by Adam Smith, we identified three types of collaborators in the author co-contributorship network: they are called specialists, versatiles, and team-players. The three types of collaborators form diverse teams and contribute to publications in various ways.

Team-players are the backbones in scientific studies. They usually contribute to the five common tasks (data analysis, performing experiments, writing articles, and contribute materials and tools). They seldom take up tasks with authorities (i.e., providing funding or project supervision). Versatiles are not that common in a team as team-players are. They are usually those who connect collaborators in a team (with edges to other collaborators) and do well in all five common tasks, with a specialty in performing experiments. They are also featured with a high level of authorities in teams. For example, study supervision and obtaining funding are their dominant tasks among the less frequent ones. Specialists are special since they usually maintain such a small population across different team sizes. Larger teams cannot eliminate them. Besides, they put themselves in a distinct position of performing collaborations. They are usually those who contribute tools, materials, and special supports. These supports can either sign their authority in a team, like providing financial support, or their blur figures, such as technical support.

These observations can help in various ways in the future: author credit assessment, team structure optimization, and candidate projects assessment guidance. Usually, author credit is given by the authors' byline position either evenly or differently (Stallings et al., 2013). These operations can be problematic sometimes when the collaboration between authors is not well assigned (e.g., Sauermann & Haeussler, 2017). Given the contributions the authors performed, their roles and credits could be given more fairly with a well-defined system of contribution scoring.

Teams vary but only a few of them succeed. And they are not simple combinations of the geniuses but of diverse roles and complementary skills (Amabile et al., 2001; Belbin, 2012). Our work might signify a way to build a scientific team with consideration of members' most frequent tasks in their earlier studies given limited resources and expense.

Similarly, the co-contributorship network may also help us to find patterns of success based on the characteristic of the team members, task division and assignment, and specialization within the teams. Thus, the funding agencies can achieve better assessment of the team structure of the proposals according to their publication history and their roles in these studies.

However, with so many aforementioned potentials, this study has several limitations. One is that this study so far takes each contribution equally, while the criteria could vary across disciplines. How to weigh different types of contributions based on specific applications could become an interesting area to study. Second, the data set of this study mainly comes from natural science, especially biology. But among the five common tasks, "contributed reagent/materials/analysis tools" is not common in social science. Extending data to social science or other fields is an important next step to follow. Third, this study only studied a collaborative team associated with a single article. It does not investigate the perspective of a researcher joining different teams and playing various roles. After author names have been disambiguated, we can address fascinating questions, such as how a scholar's career evolves based on her/his team roles in collaborations. Fourth, collaboration is becoming international. Taking nationalities, culture barriers, institutional prestige, and skill diversity into consideration, the division of labor can be further expanded to the social, behavioral, and political arena, which makes it complicated yet exciting to pursue.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their constructive suggestions for improving this article.

ENDNOTES

[1] http://journals.plos.org/plosone/s/authorship

[2] http://www.cell.com/pb/assets/raw/shared/guidelines/CRediT-taxonomy.pdf

[3] When using Equation (1) to calculate the normalized graph density, if either m or n equals one, then $m \times n = \max(m, n)$. Under this situation we decide the density is one.

REFERENCES

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, *45*(3), 425–455.

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312–313.

Amabile, T. M., Patterson, C., Mueller, J., Wojcik, T., Odomirok, P. W., Marsh, M., & Kramer, S. J. (2001). Academic-practitioner collaboration in management research: A case of cross-profession collaboration. *The Academy of Management Journal*, *44*(2), 418–431.

Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3–4), 590–614.

Batagelj, V., & Brandes, U. (2005). Efficient generation of large random networks. *Physical Review E*, *71*(3), 036113.

Becker, G. S., & Murphy, K. M. (1992). The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, *107*(4), 1137–1160.

Belbin, R. M. (2012). *Management teams: Why they succeed or fail* (3rd ed.). London: Routledge.

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology*, *57*(13), 1758–1770.

Brand, A., Allen, L., Altman, M., Hlava, M., & Scott, J. (2015). Beyond authorship: Attribution, contribution, collaboration, and credit. *Learned Publishing*, *28*(2), 151–155.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology*, *36*(11), 1224–1240.

Chompalov, I., Genuth, J., & Shrum, W. (2002). The organization of scientific collaborations. *Research Policy*, *31*(5), 749–767.

Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio- and nanoscientists' motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the*

Association for Information Science and Technology, *67*(8), 1836–1848.

Corrêa, E. A., Jr., Silva, F. N., Costa, L. D. F., & Amancio, D. R. (2017). Patterns of authors contribution in scientific manuscripts. *Journal of Informetrics*, *11*(2), 498–510.

Delfgaauw, J., Dur, R., & Souverijn, M. (in press). Team incentives, task assignment, and performance: A field experiment. *The Leadership Quarterly*.

Ding, Y., Foo, S., & Chowdhury, G. (1998). A bibliometric analysis of collaboration in the field of information retrieval. *The International Information & Library Review*, *30*(4), 367–376.

Durkheim, E. (1933). *The division of labor in society*. New York: Free Press.

Earley, P. C. (1993). East meets West meets Mideast: Further explorations of collectivistic and individualistic work groups. *Academy of Management Journal*, *36*(2), 319–348.

Erdös, P., & Rényi, A. (1959). On random graphs, I. *Publicationes Mathematicae (Debrecen)*, *6*, 290–297.

Ezzamel, M., & Willmott, H. (1998). Accounting for teamwork: A critical study of group-based systems of organizational control. *Administrative Science Quarterly*, *43*, 358–396.

Fox, M. F., & Faver, C. A. (1984). Independence and cooperation in research: The motivations and costs of collaboration. *The Journal of Higher Education*, *55*(3), 347–359.

Giles, C. L., & Councill, I. G. (2004). Who gets acknowledged: Measuring scientific contributions through automatic acknowledgment indexing. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(51), 17599–17604.

Glänzel, W. (2002). Co-authorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library Trends*, *50*(3), 461–473.

Goffman, W., & Warren, K. S. (1980). *Scientific information systems and the principle of selectivity*. New York, US: Holt-Saunders.

Guimerà, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702.

Hara, N., Solomon, P., Kim, S. L., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology*, *54*(10), 952–965.

Heffner, A. G. (1979). Authorship recognition of subordinates in collaborative research. *Social Studies of Science*, *9*(3), 377–384.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, *56*, 517–543.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. (1974). The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology*, *10*(4), 371–384.

Jabbehdari, S., & Walsh, J. P. (2017). Authorship norms and project structures in science. *Science, Technology, & Human Values*, *42*(5), 872–900.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18.

Kraut, R. E., Galegher, J., & Egido, C. (1987). Relationships and tasks in scientific research collaboration. *Human–Computer Interaction*, *3*(1), 31–58.

Larivière, V., Desrochers, N., Macaluso, B., Mongeon, P., Paul-Hus, A., & Sugimoto, C. R. (2016). Contributorship and division of labor in knowledge production. *Social Studies of Science*, *46*(3), 417–435.

Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, *66*(7), 1323–1332.

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, *11*(1), 3–15.

Leahey, E., & Reikowsky, R. C. (2008). Research specialization and collaboration patterns in sociology. *Social Studies of Science*, *38*(3), 425–440.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673–702.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273–307.

Leroy, S. (2009). Why is it so hard to do my work? The challenge of attention residue when switching between work tasks. *Organizational Behavior and Human Decision Processes*, *109*(2), 168–181.

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*(1), 31–40.

Milojević, S. (2010). Modes of collaboration in modern science: Beyond power laws and preferential attachment. *Journal of the American Society for Information Science and Technology*, *61*(7), 1410–1423.

Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, *6*(2–3), 161–180.

National Human Genome Research Institute (2019). The Human Genome Project. Retrieved from https://www.genome.gov/human-genome-project

Newman, M. E. (2004). Co-authorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5200–5205.

O'Connor, J. (1970). Growth of multiple authorship. In *Proceedings of DRTC Seminar* (Vol. 7, pp. 463–483). Bangalore, India: Documentation Research and Training Centre.

Price, D. D. S., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, *21*(11), 1011–1018.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork? *Small Group Research*, *36*(5), 555–599.

Sauermann, H., & Haeussler, C. (2017). Authorship and contribution disclosures. *Science Advances*, *3*(11), e1700404.

Science History Institute (2017). James Watson, Francis Crick, Maurice Wilkins, and Rosalind Franklin. Retrieved from https://www.sciencehistory.org/historical-profile/james-watson-francis-crick-maurice-wilkins-and-rosalind-franklin

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London, UK: Methuen & Co., Ltd.

Stallings, J., Vance, E., Yang, J., Vannier, M. W., Liang, J., Pang, L., … Wang, G. (2013). Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences*, *110*(24), 9680–9685.

Stigler, G. J. (1951). The division of labor is limited by the extent of the market. *Journal of Political Economy*, *59*(3), 185–193.

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039.

Xie, Z., Ouyang, Z., Li, J., Dong, E., & Yi, D. (2018). Modelling transition phenomena of scientific coauthorship networks. *Journal of the Association for Information Science and Technology*, *69*(2), 305–317.

Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, *60*(10), 2107–2118.

Yang, S., Wolfram, D., & Wang, F. (2017). The relationship between the author byline and contribution lists: A comparison of three general medical journals. *Scientometrics*, *110*(3), 1273–1296.

Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology*, *69*(1), 72–86.

How to cite this article: Lu C, Zhang Y, Ahn Y-Y, Ding Y, Zhang C, Ma D. Co-contributorship Network and Division of Labor in Individual Scientific Collaborations. *Journal of the Association for Information Science and Technology*. 2019;1–17. https://doi.org/10.1002/asi.24321

APPENDIX

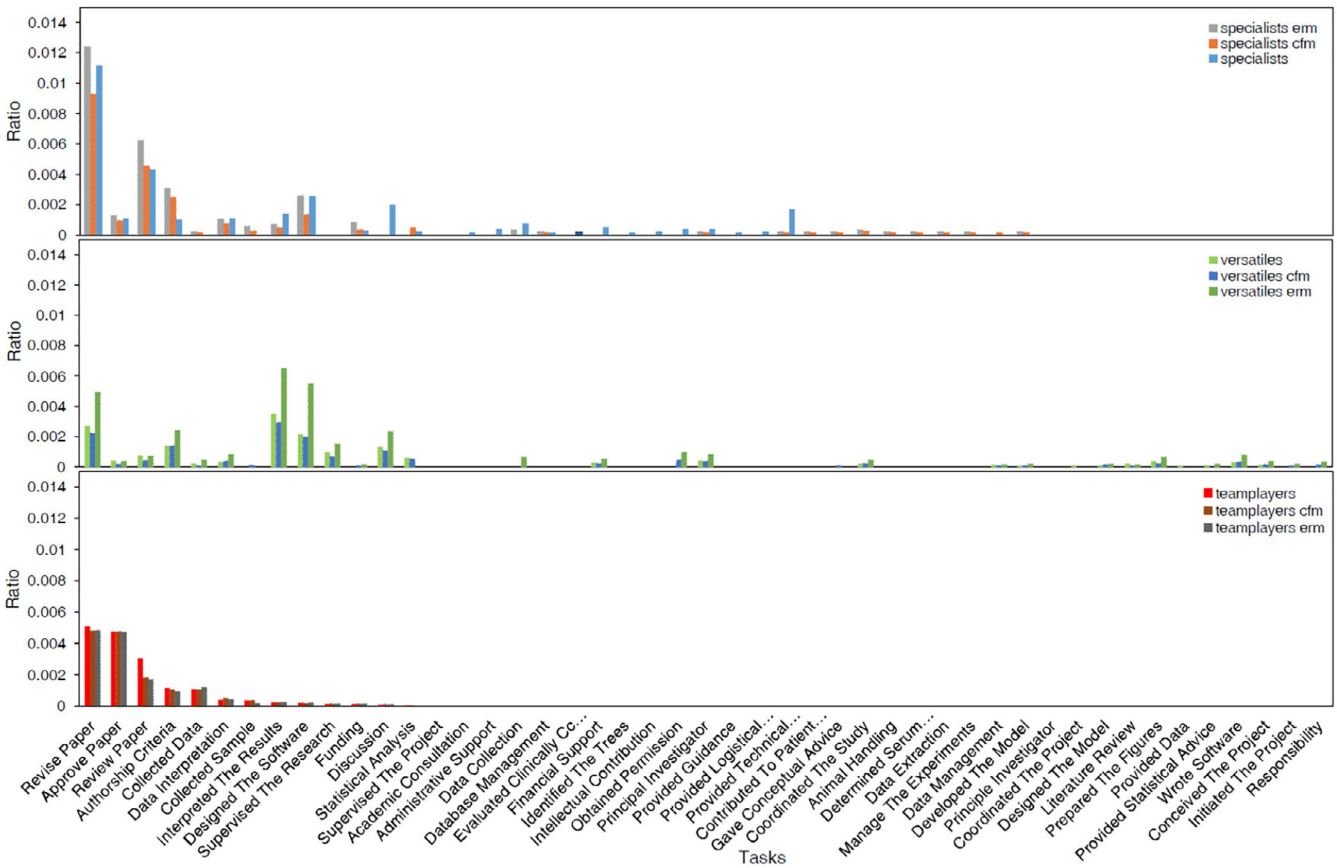


FIGURE A1 Less frequent tasks performed by different collaborators [Color figure can be viewed at wileyonlinelibrary.com]